Plant tagnology

Tamara Maes, Peter De Keukeleire and Tom Gerats

Transposable elements have been used as an effective mutagen and as a tool to clone tagged genes. Insertion of a transposable element into a gene can lead to loss- or gain-of-function, changes in expression pattern, or can have no effect on gene function at all, depending on whether the insertion took place in coding or non-coding regions of the gene. Cloning transposable elements from different plant species has made them available as a tool for the isolation of tagged genes using homologous or heterologous tagging strategies. Based on these transposons, new elements have been engineered bearing reporter genes that can be used for expression analysis of the tagged gene, or resistance genes that can be used to select for knockout insertions. While many genes have been cloned using transposon tagging following traditional forward genetics strategies, gene cloning has ceased to be the rate-limiting step in the process of determining sequence–function relations in several important plant model species. Large-scale insertion mutagenesis and identification of insertion sites following a reverse genetics strategy appears to be the best method for unravelling the biological role of the thousands of genes with unknown functions identified by genome or expressed sequence tag (EST) sequencing projects. Here we review the progress in forward tagging technologies and discuss reverse genetics strategies and their applications in different model species.

Traditional forward genetics strategies aim to clone genes that have been defined by a mutant phenotype or function. Transposon-tagged mutants can be isolated by random tagging, that is, by isolating plants with altered phenotypes from progenies resulting from self pollination of lines carrying active transposable elements. Alternatively, transposon-tagged mutants can be obtained by directed tagging, that is, by crossing individuals homozygous for a stable recessive mutation of interest with a line carrying highly active transposable elements, and isolating the few individuals with the mutant phenotype from a large F1 population. Forward gene tagging and cloning strategies have been especially successful using heterologous, low copy number lines (reviewed in Ref. 1).

Insertion flanking sequences can be easily obtained from single or low copy number lines using inverse PCR (iPCR; Ref. 2) or by thermal asymmetric interlaced (TAIL) PCR, which consists of three consecutive rounds of semi-nested PCR, performed with a set of three nested insertion-specific primers and a small, arbitrary primer, which anneals nearby in the insertion flanking sequence[3,4].

Cloning genes tagged by endogenous elements has often been complicated by the presence of many related elements in the genome. Unless the element responsible for the mutant phenotype can be identified using subfamily-specific internal probes, the identification of individual insertions is severely limited by the resolution of agarose gels in Southern analysis. Subsequent isolation of the tagged gene was equally laborious, involving the construction and screening of the size-fractionated genomic library of the insertion mutant, which explains why relatively few plant genes have been isolated using endogenous elements since the *bronze* gene of maize was isolated using the *Ac* element as a tag[5]. Recently, several PCR approaches have been designed to overcome these problems. *dTph1* tagged genes have been cloned from *Petunia* by differential screening of iPCR amplified *dTph1* flanking sequences from mutant and wild-type DNA (Ref. 6). Transposon display and amplification of insertion mutagenized sites (AIMS) are related, amplified fragment length polymorphism (AFLP©) based techniques which allow the visualization and isolation of transposon flanking sequences from high copynumber lines[7–9]. In both methods, genomic DNA is digested with restriction enzymes and provided with adaptors. One restriction enzyme recognizes a site conserved in the insertion element sequence, while a second restriction enzyme cuts nearby in the flanking genomic sequence. In Transposon Display, first round products are amplified from these restriction fragments using a chimeric primer based on the adaptor and transposon sequence, and a second adaptor primer (Fig. 1). In AIMS, these products are obtained by linear PCR with a biotinylated transposon-specific primer. First round AIMS or Transposon Display products are reamplified using a labelled, nested transposon-specific primer and an adaptor primer. The obtained transposon flanking sequences are separated on a polyacrylamide gel and visualized by autoradiography. Flanking sequences cosegregating with a mutation can be isolated from the gel, reamplified, sequenced and used as a probe to screen cDNA or genomic libraries.

The potential use of Transposon Display and AIMS is not limited to gene isolation. The high resolution and the sensitivity of the system permits detailed and simultaneous analysis of the insertion behaviour of a transposable element family, consisting of more than a hundred closely related elements. The insertion behaviour of the family can be studied in function of developmental (sampling the plant during different stages of development or sampling different organs), environmental (sampling cuttings or cell cultures derived from the same individual in different growth conditions) and genetic clues.

Reverse genetics

Reverse genetics aims to determine the function of a gene for which the sequence is known, by generating and analysing the phenotype of the corresponding knockout mutant.

Gene disruption is a powerful tool for obtaining knockout mutants that permits the assigning of biological functions to the numerous uncharacterized open reading frames (ORFs) determined by genome sequencing projects or represented in expressed sequence tag (EST) databases. Several strategies to disrupt sequenced genes or to simplify the identification of individuals in which genes have been disrupted have been proposed.

Both insertional mutagenesis and gene disruption or replacement via homologous recombination are being used for large-scale mutagenesis in *Saccharomyces cerevisiae*. Tagged yeast mutants have been generated *en masse*, either by using random transposition of the endogenous *Ty1* element[10], or by mutagenizing a yeast

1360 - 1385/99/$ – see front matter PII: S1360-1385(99)01375-8

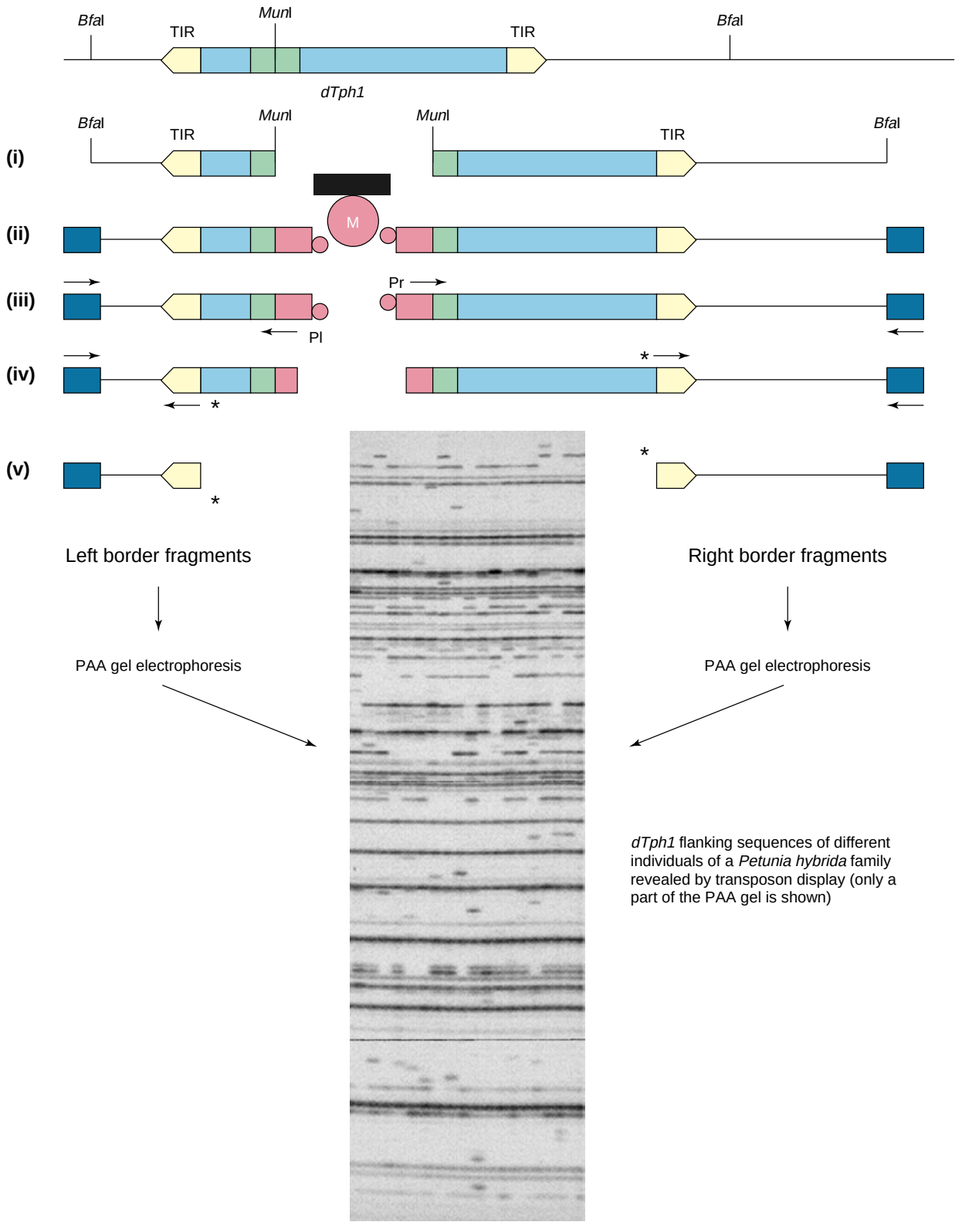

Fig. 1. Amplification and detection of transposon flanking sequences (after Ref. 7). (i) Digestion of genomic DNA with a hexacutter, that cuts the transposon, and a tetracutter (generates two TIR-containing fragments per transposon). (ii) Ligation of biotinylated hexacutter adaptor and tetracutter adaptors. Magnetic isolation of biotinylated fragments. (iii) Preamplification using tetracutter adaptor primer and one of two chimeric primers (P1 or Pr), based on the hexacutter adaptor sequence and on the transposon sequence adjacent to the hexacutter restriction site in the insertion element. (iv) 'Hot' amplification using tetracutter adaptor primer and labeled TIR primer. (v) Labelled transposon flanking fragments to be analysed by PAA gel electrophoresis. Red box, biotinylated hexacutter adaptor; blue box, tetracutter adaptor; green box, half of the imperfect internal palindromic sequence of *dTph1*; M, magnetic streptavidin beads; black box, magnet; *dTph1*, *defective transposon Petunia hybrida 1*; TIR, terminal inverted repeat; Pl, left border internal primer; Pr, right border internal primer.

Table 1. Plant tagging systems used for reverse genetics

Tag	Species	Gene trap marker		Insertion marker	Copies	Insertion site determination[a]			Refs
		Expression	Selection			Random sequence	PCR selection		
Mu	Maize	No	No	No	>100		*An1*	(1)	34
							Zag1	(1)	35
dTph1	*Petunia*	No	No	No	>100		*Nam*, *An3*	(10)	16
							Alf	(1)	
En-1	*Arabidopsis*	No	No	No	6	(15)	*F3H (TT6)*, *FLS*	(32)	36
T-DNA	*Arabidopsis*	No	No	Yes	Low		*CYP83B1-1*	(12)	37
							AKT-1	(16)	30,31
							Act-4, *Act-2*, *Act-7*		32,33
T-DNA	*Arabidopsis*	No	Yes	No	Low	*Dal1*, *SK1-B2*, *SK2-3*, *SK2-N2*, *SK33* (12)			38
T-DNA	*Arabidopsis*	No	Yes	Yes	Low	(32)			23
Ds-G	*Arabidopsis*	Yes	No	Yes	Low	*PROLIFERA* (several hundred)			39
									40

[a]Gene names, or in the case of Ref. 38, lines exhibiting a mutant phenotype. Numbers in brackets indicate the number of genes tested according to the reference. The numbers should not be used to calculate and compare the frequency of success.

library in *E. coli* using the *Tn3* element (carrying *lac*Z for expression monitoring) and subsequently replacing the endogenous yeast genes with the mutated yeast sequences via homologous recombination[11].

An alternative method involves the systematic replacement of each of the ≈6000 ORFs of *S. cerevisiae* with a synthetic PCR fragment containing a unique 20 bp tag and a kanamycin marker flanked by short sequences to facilitate homologous recombination[12]. Although homologous recombination has become technically possible in plants[13], the low efficiency of the process makes it unsuitable for large-scale functional analysis.

Transposon and T-DNA insertion mutagenesis (for a recent review on the latter see Ref. 14) appear to be the best methods for large-scale functional genomics in plants. Saturated populations of tagged mutants can be obtained using either methods and tagged genes can be identified using the insertion DNA as a tag.

Gene tagging systems used for reverse genetics

The use of insertional mutagenesis for reverse genetics is a relatively recent development in plant tagging technology.

Previously, populations containing endogenous or heterologous elements have been generated to screen for F2 individuals with mutant phenotypes, and enhancer- or gene trap-lines have been used to identify promoters or to identify genes with interesting expression patterns. While the populations generated in these efforts can also be used to perform reverse genetic screens, it should be stressed that none of the applied tagging strategies has been designed specifically for large-scale functional analysis using reverse genetics strategies. Consequently, all suffer particular shortcomings that will influence the efficiency of the recovery of knockout mutants and the subsequent phenotypic analysis. The prerequisite of an efficient reverse genetics system is that it should be possible to determine if a gene knockout has or has not been obtained, which is particularly important as gene knockouts might not lead to an easily identifiable phenotype for the majority of genes.

In the next paragraphs, several examples will be given to illustrate different tagging approaches, which are now also used for reverse genetics. A more general overview of tagging strategies used for reverse genetics in different plant species is represented in Table 1.

Endogenous tagging systems

Reverse genetic screens for insertion mutants have been set up using endogenous high-copy-number elements such as Robertson's mutator (*Mu*) in maize[15] and *dTph1* in *Petunia*[16]. The advantages of using endogenous high-copy-number elements are that:

- Large populations of insertion mutants can be obtained with ease and grown freely in the field.
- By using lines in which multiple elements are transposing, smaller populations suffice to ensure that tagged mutants will be found for every gene.

The disadvantages are that:

- Although many endogenous transposable elements like *Mu* exhibit a clear preference for insertion into genes[17], they insert in both coding and non-coding regions.
- Only a fraction of the insertions in the population will represent gene knockouts.
- Efficiency of the recovery of knockout mutants depends entirely on the ability of the system used to identify the insertion sites to distinguish knockouts from the rest of the insertion mutations.

In high-copy-number lines, each individual of the population can harbour multiple knockout mutations caused by the transposable element. The genetic linkage between a phenotype and an insertion has to be determined both in high- and single-copy lines. However, the assessment, particularly of subtle phenotypes, can be complicated substantially by the heterogeneity of high-copy-number lines and could require several generations of outcrossing to remove all but the relevant insertion.

Engineered tagging systems

Setting up a large population of individuals carrying insertions of an engineered transposon is an enormous task, as is the subsequent identification of the insertion sites and the phenotypic analysis of the isolated mutants. It is therefore important to seize all the opportunities that the use of an engineered system offers over the natural transposon tagging systems. Strategies to control frequency and timing of transposition, to monitor excision and reinsertion, and to randomize insertion positions to obtain whole genome coverage of insertion mutants have been reviewed elsewhere[18].

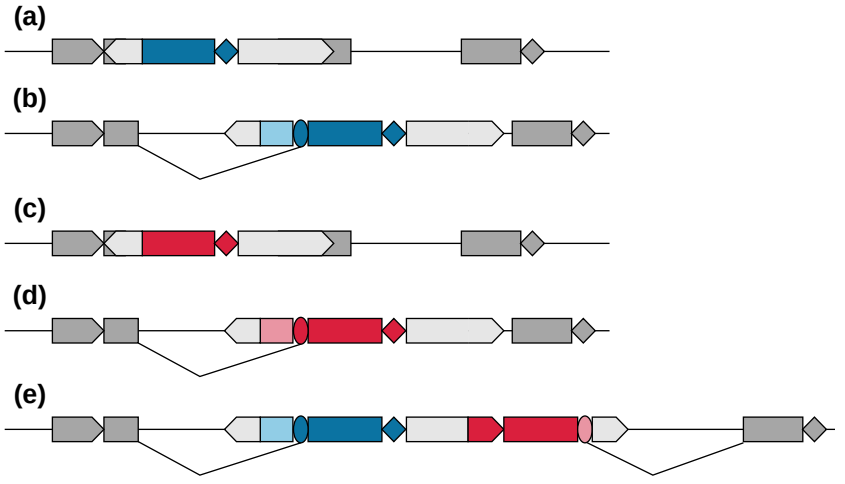

Fig. 2. Gene trapping vectors. Gene trapping vectors are designed to express their gene trap markers only when they are inserted into a gene in the correct orientation. All but the last type of vector has been used in plant insertion mutagenesis. (a) Promoterless colorimetric reporter gene is expressed when the trapping vector is inserted downstream of an active promoter[19,20]. (b) Promoterless colorimetric reporter gene is expressed when the trapping vector is inserted downstream of an active promoter and forms part of a transcript that is processed to form a functional fusion with the endogene[21]. (c) Same principle as in (a), but using a promoterless antibiotic-resistance gene as a gene trap marker. Selection of lines expressing the antibiotic marker is dependent on the activity of the promoter of the tagged gene and favours multiple insertions and/or highly expressed genes[22,23]. (d) The same principle as in (b), using a promoterless antibiotic-resistance gene as a gene trap marker[24]. (e) This vector combines the features of (b) with an approach designed to select gene trapping events without the bias for insertions presented by selection-based trapping vectors that depend on endogene expression like (c) and (d). A constitutive promoter drives the expression of the antibiotic-resistance-marker gene, which ends in a consensus splice donor site. Expression of antibiotic resistance can be achieved when the splice donor site is joined to an endogenous splice acceptor site and when a polyadenylation signal is provided by the endogenous sequence[38]. Blue box, colorimetric reporter gene; light blue box, intron preceding the colorimetric reporter gene; blue or red oval, splice acceptor site preceding marker; red box, antibiotic resistance gene; pink oval, splice donor site following marker; pink box, intron preceding antibiotic resistance gene; dark grey box, exon of the tagged gene; dark grey arrow, promoter of the tagged gene; red arrow, constitutive promoter of the antibiotic-resistance gene; light grey arrows, gene trapping vector ends; diamonds, polyadenylation signals; black lines joining different boxes indicate the formation of splicing products in the corresponding mRNAs.

Gene trapping vectors that enable screening or selection for insertions into genes have been engineered for use in mammals, *Drosophila* and plants. Gene trapping can be achieved by making the expression of a marker gene in the tagging vector (either transposon or T-DNA based) dependent on insertion into a gene.

This dependency can be created in various ways. A first option is to include a promoterless marker gene in the tagging vector, which will only be transcribed when inserted downstream of an active endogenous plant promoter. In plants, various T-DNA and transposon based vectors have been designed that incorporate a promoterless *uidA* reporter gene encoding for β-GUS activity as a tagging marker[19,20] (Fig. 2a).

Strictly speaking, tagging vectors based on this principle do not allow the detection of insertions in genes but of the insertions downstream of a functional promoter. A more stringent screen for gene insertions can be obtained by making the expression of the marker in the tagging vector dependent not only on the activity of an endogenous promoter but also on transcript processing (splicing) and the formation of translational fusions. This was obtained by introducing an intron and splice acceptor sites in front of the *uid*A reporter gene in a *Ds*-based vector[21]. Transcripts containing the *uid*A gene need to be processed by splicing out the intron, and the *uid*A coding sequence needs to be joined in-frame to part of the sequence of an endogene to confer β-GUS activity (Fig. 2b).

A clear advantage of these gene trapping systems is that the expression pattern of the tagged gene can be studied in detail by analysing the GUS staining pattern during the development of the plant. Such detailed knowledge of the expression pattern can be very helpful in the subsequent phenotypic analysis of the homozygous insertion mutants. Their disadvantage is that there is neither selection against insertions outside genes, nor against insertions in which the *uid*A reporter gene is in the opposite orientation relative to transcription of a tagged gene.

Other gene trapping vectors have been designed to select for, rather than to screen for, lines in which a gene has been tagged. This can be achieved by making the expression of a selection marker (rather than a colorimetric marker) in the tagging vector dependent on insertion into a gene.

The idea of including promoterless selection markers, that would only be activated after formation of functional transcriptional- or translational-fusions with expressed endogenous genes near the T-DNA borders, was described as early as 1989 by Konzc *et al*., but was not considered to be satisfactory[22]. A T-DNA construct containing a promoterless selectable marker [*aph*(3′)II], conferring resistance to kanamycin, was used to transform plant tissue (Fig. 2c). Selection on kanamycin, however, inherently biases the procedure towards recovery of calli that exhibit higher *aph*(3′)II expression levels, and favours T-DNA integration into genes that are (highly) expressed during the selection procedure and/or multiple T-DNA integrations. These problems were circumvented by including a hygromycin-resistance gene for the primary selection of transgenics in the T-DNA and using the *aph*(3′)II gene only as a secondary reporter, albeit at the cost of abandoning the idea to select directly for transformants in which genes had been tagged[22,23].

The approach mentioned above exemplifies a real gene trapping system in plants, in the sense that it allows for selection of, rather than screening for, insertions into genes. However, it suffers from two important drawbacks: it lacks the versatile expression analysis facilities provided by GUS-based promoter or gene trapping systems, and the isolation of tagged individuals based on the expression of the promoterless resistance gene is dependent on the expression state of the tagged gene(s).

In the previous examples, the expression of the marker gene in the gene trapping vector depends on the activity of the endogenous promoter. However, selection or screening for gene-specific insertions can also be accomplished independently of endogenous promoter activity, by making the expression of a marker in the gene trapping vector dependent on proper mRNA synthesis and termination. Both options were incorporated in a beautiful gene trapping system designed to perform high throughput mutagenesis of mouse embryonic stem cells[24]. A retroviral gene trapping vector

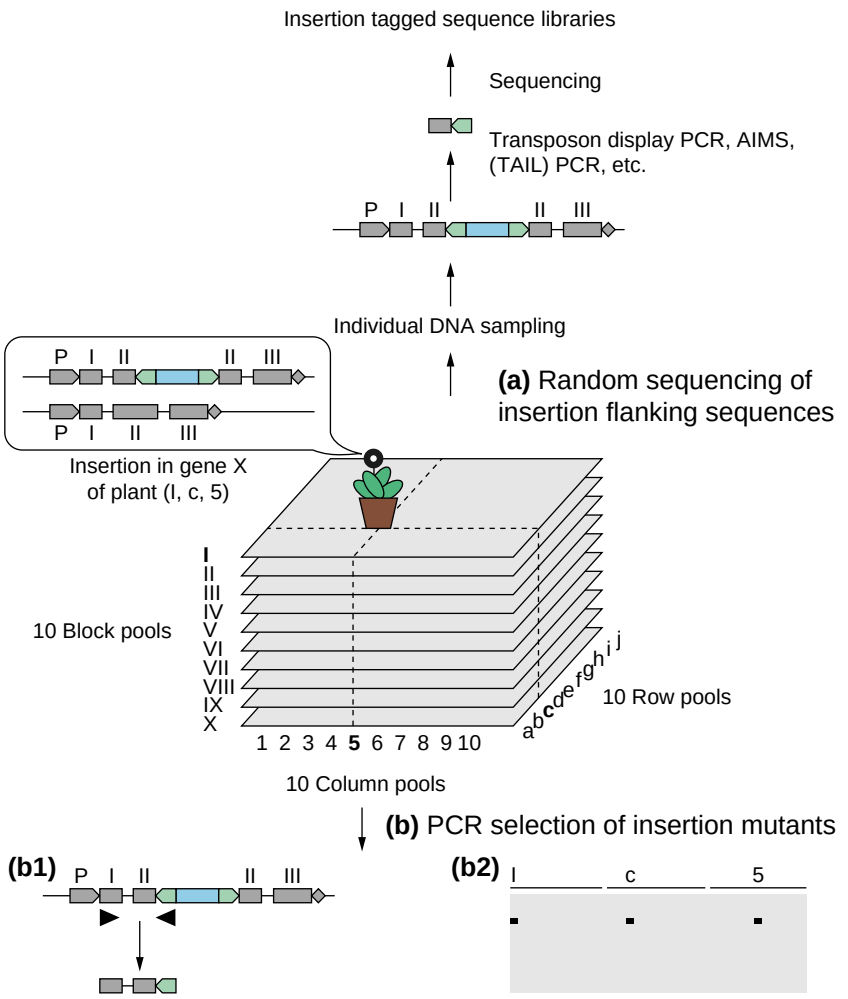

Fig. 3. Identification of insertion sites in reverse genetics; the two main roads leading to the identification of insertion. A population of 1000 plants harbours an individual (I, c, 5), which carries an insertion in gene X. Insertions in gene X can be identified in two ways: by (a) random sequencing of insertion flanking sequences or by (b) PCR selection of insertion mutants. Random sequencing is achieved by preparing separate genomic DNA samples from every individual and by systematically amplifying the insertion flanking sequences represented in these samples using one of the many available PCR-based methods (see text). The obtained fragments are sequenced and catalogued in insertion tagged sequence databases. PCR selection of insertion mutants can be used to screen for insertions in specific genes. The screens are not done on individuals but on pools of plants to reduce the number of PCR reactions. The population is organized in a set of pools that represent the population in a three-dimensional (3D) or pyramidal fashion. For example, the 1000 plants of this population were organized in a 3D consisting of ten blocks, rows and columns consisting of 100 plants. The DNA samples prepared from the 30 pools are subjected to a PCR reaction using a gene-X-specific primer and an insertion-specific primer. Pools containing DNA from plant (I, c, 5), give rise to the amplification of a transposon flanking fragment (b1) that cross-hybridizes with a gene-X-specific probe (b2). Abbreviations: (TAIL) PCR, thermal asymmetric interlaced PCR; AIMS, amplification of insertion mutagenized sites; P, endogenous promoter; I, II, II, exons of the interrupted gene; green arrows, termini of the insertion element; blue box, internal part of the insertion element.

containing two important cassettes was used. The expression analysis cassette contains a splice acceptor and an internal ribosome entry site in front of a selectable, colorimetric marker (β-galactosidase–*npt*II fusion gene) and a polyadenylation signal. The expression cassette allows a detailed analysis of the expression profile of the tagged gene. The selection cassette contains a puromycin-resistance gene, which is expressed under control of a constitutive promoter in the retroviral construct. The puromycin gene lacks a polyadenylation sequence and terminates in a consensus splice donor site. Puromycin resistance can be obtained when the splice donor site is joined with a downstream splice acceptor site of an endogenous exon and when a polyadenylation site is provided by the host genome (i.e. when the element has been inserted into a gene) (Fig. 2e).

Applying the same strategy to plants could significantly reduce the number of lines that need to be generated and analysed, without having to sacrifice the advantages offered by expression analysis facilities.

The engineered gene trap systems described above contain reporter or selection genes that will be expressed when a gene has been hit. To prevent non-gene-specific insertions being co-isolated with gene hits, and to exclude the possibility that the observed expression pattern represents the sum of the expression pattern of several tagged genes, these approaches are necessarily limited to single-copy insertion libraries.

Determining the insertion sites: random sequencing versus site-selected approaches

In the previous section, different gene tagging systems used to generate a large population of insertion mutants for reverse genetics have been illustrated. To identify tagged genes in such mutagenized populations, two main options are currently available: to randomly amplify and sequence insertion flanking sequences, or to specifically screen for insertions in genes of interest (Fig. 3a). Methods have been developed that make these two options feasible, both in single- and high-copy-number lines.

Insertion flanking sequences can be amplified from genomic DNA using iPCR, Transposon Display, AIMS, (TAIL) PCR or related techniques[2–4,7,9]. Alternatively, insertion flanking sequences can be amplified from first-strand cDNA using 5′ or 3′ rapid amplification of cDNA ends (RACE; Ref. 24). To determine insertion flanking sequences from single-copy lines, amplified products can be sequenced directly. In high-copy-number lines, the amplification products derived from different insertion sites are resolved on sequencing gels, isolated, re-amplified and sequenced individually.

To screen for insertions into specific genes, a PCR-based strategy, described originally for site-specific selection of P element insertions[25,26], has been applied to different tagging systems and different model plants like *Petunia*, *Arabidopsis* and maize. The philosophy of the method is that the insertion of a known element into a gene can be identified by the amplification of a PCR product using a gene-specific and an insertion-specific primer (Fig. 3b1). The sensitivity of the PCR technique, especially after hybridization of the PCR products with a gene-specific probe, allows the easy detection of a single gene hit within a pool of hundreds or thousands of individuals. Screenings of pools of

insertion mutants are generally organized in a three-dimensional or in a pyramidal setup, to allow easy identification of tagged individuals[27] (Fig. 3).

Which one of the two methods, site-selected PCR screening or random sequencing, will be more appropriate depends on:

- The tagging system.
- Whether the entire genome sequence of the model organism is available.
- The speed at which mutants can be identified by either method.

Systematic sequencing of insertion tags will undoubtedly generate large databases of flanking sequences, which can readily be used to identify the tagged gene by sequence comparison if the whole genome sequence is available.

As long as whole genome catalogues of insertion mutants are not available, specific PCR-based screens will be faster and thus desirable for the recovery of mutants in, for example, a set of genes differentially expressed under certain conditions, or of members of a gene family. However, to meet large-scale functional genomic requirements and to compete with random sequencing (to find mutants for 'all' of the genes in a given plant species in a time- and cost-effective manner), new PCR screening methods, allowing parallel identification of many insertions in different genes, need to be developed.

Future prospects

Expression profiles obtained by the analysis of gene trap or reporter lines described above, or by whole genome expression analysis on microarrays (glass slides displaying thousands of spatially addressable, gene-specific cDNA or oligonucleotide targets that can be hybridized with labelled first-strand cDNA)[28], might reveal joint expression patterns for genes that are involved in the same process. If the whole genome sequence is available, *in silico* comparisons of the upstream regions of genes expressed in the same tissues, developmental stages or specific conditions might lead to the identification of common motives that control this expression specificity. Algorithms to deal with such questions have already been developed and applied to identify promoter elements of concertedly expressed genes in yeast[29]. The availability of expression profile catalogues will allow the genetic and molecular interactions of genes thought to be involved in a certain pathway to be studied in a much more directed manner.

Strong evidence for the biological function of a gene can come from the phenotype of the corresponding knockout mutant. Results obtained from single-gene-knockout experiments in many different eukaryote systems, suggest that there is a high level of redundancy in gene function. The question of redundancy can be addressed systematically once knockouts for all the members of a gene family have been identified, although we need to keep in mind that functional redundancy might also occur between non-homologous genes. Nevertheless, the function of the majority of genes will not be revealed by the analysis of knockout phenotypes, because laboratory conditions are not assaying for that function. A good example is the *Arabidopsis* T-DNA insertion mutant, isolated for the root-specific K^+ channel protein *AKT1* using a PCR-based reverse genetic screen. Plants homozygous for the *akt1-1* mutation do not exhibit any obvious mutant phenotype on several different nutrient media. However, growth of the *akt1-1* mutant is significantly inhibited on media containing low concentrations of K^+, in the presence of NH_4^+ (Refs 30, 31). Another assay reveals the subtle but real phenotype of actin T-DNA insertion mutants. Individuals homozygous for one of the T-DNA tagged actin alleles, *act2-1*, *act4-1* or *act7-1*, do not have a clear mutant phenotype. However, the observation that the frequency of these alleles consistently decreases when populations segregating for these alleles are being analysed during multiple generations, illustrates that these mutations lead to reduced plant fitness[32,33].

In yeast, various strategies have been developed to identify genes that either confer resistance or sensitivity to different growth conditions. A large population of deletion mutants, in which each of the 6000 yeast genes is replaced by a kanamycin resistance gene and a unique 20 bp tag (which can be amplified using a set of primers complementary to the tag flanking sequence in all the mutants), can be grown in conditions that are toxic for the majority of the cells. Gene knockouts in surviving strains can be readily identified by hybridization of their unique 20 bp tags to a microarray displaying all the different 20 bp tags in a spatially addressable fashion[12]. In another approach, termed genomic fingerprinting, a saturated population of transposon *Ty* insertion mutants, is interrogated with the site-selected PCR screening method for insertion mutants. A characteristic genomic fingerprint of transposon flanking sequences is encountered for every gene-specific primer used. Propagating the population using various growth conditions leaves the genomic fingerprint of a specific gene intact when it is not essential in these conditions, but will quickly lead to the depletion of bands associated with its knockout mutants when it is required for survival in those conditions[10].

These strategies exploit the small genome size and short generation time of yeast, the easily controlled growth conditions in cell cultures and the ability of some yeast strains to proliferate as haploid cells.

While several phenotypic screens can be carried out on large insertion mutant populations in plants (most of these in the F2 generation), it is not feasible to submit these plant collections to a large set of resistance or independence screens. A single whole-genome multigenerational screen for knockout mutants displaying reduced fitness in a certain growth condition, as described above for yeast, would require warehouses. Screens might be done on plant cell cultures derived from a (sub)population of knockout mutants or on a cell culture derived from haploid plants and submitted to saturation mutagenesis. At the plant level, however, it seems more reasonable to develop strategies that divide a population of insertion mutants into subgroups that need to be examined for changed aspects in several specific conditions.

The different fields of plant functional genomics are converging to form an extremely powerful instrument. Sequence information obtained in genome sequencing projects could give the first clue about a possible gene function if it shows homology with a gene of known function. Expression analysis by DNA microarray hybridization, representative differential analysis (RDA), differential display or cDNA AFLP can be used to identify genes involved in a defined process. Finally, sequence and expression data can be combined to design appropriate phenotypic assays for the knockout mutants obtained from large-scale insertion mutagenesis projects, which will reveal the role(s) played by the gene in question.

The great potential of reverse genetics in the era of functional genomics could suggest that the time of forward genetics has definitely ended. This is not true: catalogues of insertion mutants generated in reverse genetics strategies can significantly increase the efficiency of directed transposon tagging and map-based cloning. Mapping a mutation of interest is the first step towards cloning a gene identified by a mutant phenotype. Transposon insertions, which are linked to the mutation, can be used as a launchpad for local mutagenesis in directed tagging. An alternative approach will emerge when both the genome sequence and the genome covering catalogues of insertion mutants are available for a plant species. After roughly mapping the mutation, the sequence of the closest linked markers might define a stretch of DNA containing around a hundred genes. The sequences of these genes can indicate which genes in this region are more likely to be affected in the mutant. The lines containing knockout insertion mutations for these genes can be retrieved from catalogued insertion mutant libraries. The knockout

phenotypes can then be examined in the homozygotes, or the insertion mutations subjected to allelism tests with the originally mapped mutation. As both forward and reverse genetics branches of insertion mutagenesis come of age, it will become clear that there is no fundamental difference between the two approaches. The only arbitrary difference will lie in the order of what is done first: the determination of the insertion site or the determination of the phenotype.

Conclusions

Functional genomics is an emerging approach, which combines genome sequence analysis, genome-wide mRNA or protein expression monitoring, genetic screens, and mutant isolation to form a powerful tool that allows the study of gene function on a genomic scale. Functional genomics will lead to an explosion in our understanding of the function of thousands of individual genes and the position of these genes in genetic networks. This integrated approach might, however, only become available to many plant scientists if the outputs of the different (expensive) technological platforms involved become generally accessible.

References

1 Osborne, B.I. and Baker, B. (1995) Movers and shakers: maize transposons as tools for analyzing other plant genomes, *Curr. Opin. Cell. Biol.* 7, 406–413

2 Ochman, H., Gerber, A.S. and Hartl, D.L. (1988) Genetic applications of an inverse polymerase chain reaction, *Genetics* 120, 621–623

3 Liu, Y.G. and Whittier, R.F. (1995) Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking, *Genomics* 25, 674–681

4 Smith, D. *et al.* (1996) Characterization and mapping of *Ds*-GUS-T-DNA lines for targeted insertional mutagenesis, *Plant J*. 10, 721–732

5 Fedoroff, N., Furtek, D. and Nelson, O., Jr (1984) Cloning of the *bronze* locus in maize by a simple and generizable procedure using transposable controlling element *Ac*, *Proc. Natl. Acad. Sci. U. S. A.* 81, 3825–3829

6 Souer, E. *et al.* (1995) A general method to isolate genes tagged by a high copy number transposable element, *Plant J.* 7, 677–685

7 Van den Broeck, D. *et al*. (1998) *Transposon Display* identifies individual transposable elements in high copy number elements, *Plant J*. 13, 121–129

8 Vos, P. *et al*. (1995) AFLP: a new technique for DNA fingerprinting, *Nucleic Acids Res.* 23, 4407–4414

9 Frey, M., Stettner, C. and Gierl, A. (1998) A general method for gene isolation in tagging approaches: amplification of insertion mutagenised sites (AIMS), *Plant J*. 13, 717–721

10 Smith, V., Botstein, D. and Brown, P.O. (1995) Genetic footprinting: a genomic strategy for determining a gene's function given its sequence. *Proc. Natl. Acad. Sci. U. S. A.* 92, 6479–6483

11 Burns, N. *et al*. (1994) Large-scale analysis of gene expression, protein localization and gene disruption in *Saccharomyces cerevisiae*, *Genes Dev*. 8, 1087–1105

12 Shoemaker, D.D. *et al*. (1997) Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy, *Nat. Genet*. 14, 450–456

13 Kempin, S.A. *et al.* (1997) Targeted gene disruption in *Arabidopsis*, *Nature* 389, 802–803

14 Azpiroz-Leehan, R. and Feldmann, K.A. (1997) T-DNA insertion mutagenesis in *Arabidopsis*: going back and forth, *Trends Genet*. 13, 152–156

15 Das, L. and Martienssen, R. (1995) Site selected transposon mutagenesis at the *hcf106* locus in maize, *Plant Cell* 7, 287–294

16 Koes, R. *et al.* (1995) Targeted gene inactivation in petunia by PCR-based selection of transposon insertion mutants, *Proc. Natl. Acad. Sci. U. S. A.* 92, 8149–8153

17 Cresse, A.D. *et al.* (1995) Mu1-related transposable elements of maize preferentially insert into low copy number DNA, *Genetics* 140, 315–324

18 Sundaresan, V. (1996) Horizontal spread of transposon mutagenesis: new uses for old elements, *Trends Plant Sci.* 6, 184–190

19 Fedoroff, N.V. and Smith, D.L. (1993) A versatile system for detecting transposition in *Arabidopsis*, *Plant J.* 3, 273–289

20 Lindsey, K. *et al.* (1993) Tagging genomic sequences that direct transgene expression by activation of a promoter trap in plants, *Transgenic Res.* 2, 33–47

21 Sundaresan, V. *et al.* (1995) Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements, *Genes Dev*. 9, 1797–1810

22 Koncz, C. *et al.* (1989) High-frequency T-DNA-mediated gene tagging in plants, *Proc. Natl. Acad. Sci. U. S. A.* 86, 8467–8471

23 Mathur, J. *et al.* (1998) Gene identification with sequenced T-DNA tags generated by transformation of *Arabidopsis* cell suspension, *Plant J.* 13, 701–716

24 Zambrowicz, B.P. *et al.* (1998) Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells, *Nature* 392, 608–611

25 Ballinger, D.G. and Benzer, S. (1989) Targeted gene mutation in *Drosophila*, *Proc. Natl. Acad. Sci. U. S. A.* 86, 9402–9406

26 Kaiser, K. and Goodwin, S.F. (1990) 'Site selected' transposon mutagenesis of *Drosophila*, *Proc. Natl. Acad. Sci. U. S. A.* 87, 1686–1690

27 Zwaal, R.R. *et al.* (1993) Target-selected gene inactivation in *Caenorhabditis elegans* using a frozen transposon insertion mutant bank, *Proc. Natl. Acad. Sci. U. S. A.* 90, 7431–7435

28 Schena, M. *et al.* (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray, *Science* 270, 467–470

29 Roth, F.P. *et al.* (1998) Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole genome mRNA quantitation, *Nat. Biotechnol.* 16, 939–945

30 Krysan, P.J. *et al*. (1996) Identification of transferred DNA insertions within *Arabidopsis* genes involved in signal transduction and ion transport, *Proc. Natl. Acad. Sci. U. S. A.* 93, 8145–8150

31 Hirsch, R.E. *et al.* (1998) A role for the AKT1 potassium channel in plant nutrition, *Science* 280, 918–921

32 McKinney, E.C. *et al.* (1995) Sequence-based identification of T-DNA insertion mutations in *Arabidopsis*: actin mutants *act2*-1 and *act4*-1, *Genetics* 149, 663–675

33 Gilliland, L.U. *et al*. (1998) Detection of deleterious genotypes in multigenerational studies. I. Disruptions in individual *Arabidopsis* actin genes, *Genetics* 149, 717–725

34 Bensen, R.J. *et al*. (1995) Cloning and characterization of the maize An1 gene, *Plant Cell* 7, 75–84

35 Mena, M. *et al.* (1996) Diversification of C-function activity in maize flower development, *Science* 274, 1537–1540

36 Wisman, E. *et al.* (1998) Knock-out mutants from an En-1 mutagenized *Arabidopsis thaliana* population generate phenylpropanoid biosynthesis phenotypes, *Proc. Natl. Acad. Sci. U. S. A.* 95, 12432–12437

37 Winkler, R.G. *et al.* (1998) Systematic reverse genetics of transfer-DNA-tagged lines of *Arabidopsis*. Isolation of mutations in the cytochrome p450 gene superfamily, *Plant Physiol*. 118, 743–750

38 Babiychuk, E. *et al*. (1997) Efficient gene tagging in *Arabidopsis thaliana* using a gene trap approach, *Proc. Natl. Acad. Sci. U. S. A.* 94, 12722–12727

39 Springer, P.S. *et al*. (1995) Gene trap tagging of *PROLIFERA*, an essential MCM2-3-5-like gene in *Arabidopsis*, *Science* 268, 877–880

40 Martienssen, R.A. (1998) Functional genomics: probing plant gene function and expression with transposons, *Proc. Natl. Acad. Sci. U. S. A.* 95, 2021–2026

Tamara Maes* is at the Departamento de Genética Molecular, Instituto de Biología Molecular de Barcelona (IBMB), Centro de Investigación y Desarrollo (CID), Consejo Superior de Investigaciónes Científicas (CSIC), Barcelona, Spain. Peter De Keukeleire and Tom Gerats are at the Flemish Institute of Biotechnology (VIB), Laboratorium of Genetics, University of Ghent, Belgium.

*Author for correspondence (tel +34 93 4006100; fax +34 93 2045904; e-mail tamgmr@cid.csic.es).